UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/493,748	01/28/2000	Kenji C. Obata	MSFT112958	2621

26389   7590   07/27/2004

CHRISTENSEN, O'CONNOR, JOHNSON, KINDNESS, PLLC
1420 FIFTH AVENUE
SUITE 2800
SEATTLE, WA 98101-2347

EXAMINER
NGUYEN, MAIKHANH

ART UNIT	PAPER NUMBER
2176	

DATE MAILED: 07/27/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO 90C (Rev 10/03)

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _20 April 2004_.

2a)☒ This action is **FINAL.**   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-20_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-20___ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

DETAILED ACTION

1.   This action is responsive to communications: Amendment A filed 04/20/2004 to the

original application filed 01/28/2000.

2.   Claims 1-20 are currently pending in this application. Claims 1-2, 5, 10, 12, and 17 have

been amended by Applicant. Claims 1 and 10 are independent claims.

Rejections - 35 USC § 102

3.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the

basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless -

(e) the invention was described in (1) an application for patent, published under section 122(b),by another filed in the United States before the invention by the applicant for patent or (2)a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a)shall have the effects for the purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2)of such treaty in the English language; or " (Emphasis added.)

Claims 1- 4, 6, and 10-20 remain rejected under 35 U.S.C. 102(e) as being anticipated by

Chakrabarti et al. (U.S. 6,418,433 – filed 01/1999).

As to independent claim 1, Chakrabarti teaches *(col.5, lines 14-60)* a

computer-implemented method for selectively accessing a document during a current crawl *(a*

user can search ...Web pages of interest) of a server computer *(Web server)*, the document being

identified by a document address specification *(a Web page URL)*, the document having been

retrieved during a previous crawl *(new page/old page)*, the method comprising:

- determining whether to access the document during the current crawl with the aid of a

probabilistic model that is based on the probability that the document has changed since the

previous crawl *(e.g., evaluate for potential changes in the old pages that might have occurred since last time the old pages were considered by the system; col.8, lines 53-67)*;

    - accessing the document if the determination produces an instruction indicative that the document at the document address specification should be accessed during the current crawl *(e.g., retrieve only the modified portions ...the portions that the associated Web server indicates have changed since the last time the page was considered by the system ...if the page is an old page that has been determine to have changed ...retrieve the entire page from the associated Web server; col.9, lines 45-63)*.

    As to dependent claim 2, Chakrabarti teaches computing a probability that the document has changed since the document was retrieved during the previous crawl *(e.g., a check sum representative of the page's content is computed; col.9, lines 56-63)*.

    As to dependent claim 3, Chakrabarti teaches selecting an active probability indicative of a proportion of documents in a plurality of documents that are changing at various change rates, the plurality of documents including the document *(e.g., indicating the date and time of the Web page was last modified; col.5, lines 47-60)*; training the active probability to reflect experience with the document during a plurality of previous crawls *(e.g., the topic itself can be defined by a user or by considering the seed set using the topic analyzer, including an associated classifier trainer; col.6, lines 7-15)* ; and using the trained active probability to compute the probability that the document has changed *(e.g., a check sum representative of the page's content is computed; col.9, lines 56-63)*.

    As to dependent claim 4, Chakrabarti teaches selecting the probability that the document has changed from the previous crawl as the active probability in the current crawl; and repeating

the method of Claim 3 for the current crawl (e.g., indicating the date and time when the page was

initially found ... *indicating the date and time of the Web page was last modified; col.5, lines 47-*

60).

As to dependent claim 6, Chakrabarti teaches training a document probability

distribution corresponding to the document address specification to reflect an experience with

the document during a plurality of previous crawls, the document probability distribution

including a plurality of probabilities *(e.g., the topic itself can be defined by a user or by*

considering the seed set using the topic analyzer, including an associated classifier trainer;

col.6, lines 7-15); determining from the document probability distribution a probability that the

document has changed *(e.g., evaluate for potential changes in the old pages that might have*

occurred since last time the old pages were considered by the system; col.8, lines 53-67); and

making a determination of whether to access the document in a current crawl based on the

probability that the document has changed *(e.g., retrieve only the modified portions ...the*

portions that the associated Web server indicates have changed since the last time the page was

considered by the system ...if the page is an old page that has been determine to have changed

...retrieve the entire page from the associated Web server; col.9, lines 45-63).

As to independent claim 10, Chakrabarti teaches *(col.5, lines 14-60)* a

computer-readable medium having computer-executable instructions for retrieving one document

in a plurality of documents *(a user can search ... Web pages of interest)* from a remote server

(the Web server), which when executed comprise:

- maintaining historical information associated with changes to the one document *(the Web page table ...indicating the date and time the Web page was last modified by the provider of the content of the page; col.5, lines 29-60)*;

- initiating a crawl procedure for retrieving particular documents in the plurality of documents *(collectively downloaded Web pages that are related to a limited number of predefined topics; col.5, line 61-col.6, line 15/ crawl database 30 to retrieve a list of relevant Web pages; col.6, lines 35-51)*; and

- determining whether to access the one document from the remote server based on a probabilistic analysis of the historical information associated with the changes to the one document *(e.g., retrieve only the modified portions ...the portions that the associated Web server indicates have changed since the last time the page was considered by the system ...if the page is an old page that has been determine to have changed ...retrieve the entire page from the associated Web server; col.9, lines 45-63)*

As to dependent claim 11, Chakrabarti teaches if the determination to access the one document is positive, identifying the one document for retrieval during the crawl procedure; and attempting to retrieve all documents identified for retrieval during the crawl procedure *(e.g., if the page is old page that has been determine to have changed ...if the page is determined to be a new page ... to retrieve the entire page from the associated Web server; col.9, lines 56-63)*,

As to dependent claim 12, Chakrabarti teaches computing a probability that the one document has changed since the one document was last retrieved from the remote server *(e.g., a check sum representative of the page's content is computed; col.9, lines 56-63)*.

As to dependent claim 13, Chakrabarti teaches beginning with a probability that a pre-defined proportion of documents in the plurality of documents has changed, training the probability that the pre-defined proportion of documents has changed using the historical information associated with the one document to achieve the probability that the one document has changed *(See Crawl table entry Fig.1 and associated text in col.5, lines 14-60).*

As to dependent claim 14, Chakrabarti teaches making a random decision to retrieve the one document wherein the random decision is biased by the probability that the one document has changed *(e.g., Notionally associated with each category is a many-side coin. Each face of the coin represents a words; the probability that the face come ups corresponds with the probability that the corresponding word occurs in a document of this particular category ...until the length is reached; col.7, lines 20-27).*

As to dependent claim 15, Chakrabarti teaches the random decision is further biased by a synchronization level configured to influence the random decision based on a predetermined degree of tolerance for not retrieving the one document if the document is likely to have changed *(e.g., Notionally associated with each category is a many-side coin. Each face of the coin represents a words; the probability that the face come ups corresponds with the probability that the corresponding word occurs in a document of this particular category ...until the length is reached; col.7, lines 20-27).*

As to dependent claim 16, Chakrabarti teaches the random decision is made by a software routine adapted to simulate a flip of a coin (e.g., *each face of the coin represents a words; the probability that the face come ups corresponds with the probability that the corresponding word occurs in a document of this particular category; col.7, lines 18-27).*

As to dependent claim 17, Chakrabarti teaches the historical information associated with changes to the one document includes a time stamp for the one document, the time stamp being indicative of the time that the one document was last modified when the one document was last retrieved from the remote server; and the probabilistic analysis includes a comparison of the time stamp included in the historical information with another time stamp associated with the one document stored on the remote server *(e.g., three time stamp fields are provided ... indicating the date and time the Web page was last modified; col.5, lines 47-60).*

As to dependent claim 18, Chakrabarti teaches if the time stamp included in the historical information does not match the other time stamp associated with the one document stored on the remote server, identifying the one document for retrieval during the crawl procedure *(e.g., a relevance field indicates the relevance of the Web page; col.5, lines 47-60).*

As to dependent claim 19, Chakrabarti teaches the historical information associated with changes to the one document includes a hash value associated with the one document, the hash value being a representation of the one document; and the probability analysis includes a comparison of the hash value included in the historical information with another hash value calculated from information retrieved from the one document stored on the remote server *(e.g., 64 bit hash of the URL ... determine whether any changes in the Web page have occurred; col.5, lines 29-46).*

As to dependent claim 20, Chakrabarti teaches if the hash value included in the historical information does not match the other hash value associated with the one document stored on the remote server, identifying the one document for retrieval during the crawl procedure *(e.g., a relevance field indicates the relevance of the Web page; col.5, lines 47-60).*

Claim Rejections - 35 USC § 103

4.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 5 and 7-9 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Chakrabarti et al.

As to dependent claim 5, Chakrabarti does not explicitly teach "training the active

probability includes multiplying the active probability indicative of a change in the document by

a training probability calculated using a probabilistic model."

However, Chakrabarti suggests "the probability that it was generated ... is computed

using Bayes Rule; col.7, lines 3-17".

It would have been obvious to one of ordinary skill in the art at the time of the invention

was made to apply Chakrabart's teaching for implementing the feature above in order to

determine the relevance of a document and to build a comprehensive topic – specific library for

the benefit of specific users.

As to dependent claim 7, Chakrabarti does not explicitly teach "calculating, based on

the experience with the document during a plurality of previous crawls, a discrete random

variable distribution that includes a plurality of training probabilities; multiplying each

probability in the document probability distribution by a corresponding training probability from

the discrete random variable distribution."

However, Chakrabarti suggests "the priority of a document can not only can be determined by determining its relevance, but also by determining its 'popularity', a measure of the quality of the document ... to sum to unity; col.7, lines 18-65).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to apply Chakrabart's teaching for implementing the feature above in order to build a comprehensive topic – specific library for the benefit of specific users.

As to dependent claim 8, Chakrabarti does not explicitly teach "the training probabilities are calculated using a Poisson process, the Poisson process including a Poisson equation $(e^{\wedge}(-r*dt))$ and a complementary Poisson equation $(1-e^{\wedge}(-r*dt))$."

However, Chakrabarti suggests "the probability that it was generated ... is computed using Bayes Rule; col.7, lines 3-17".

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to apply Chakrabart's teaching for implementing the feature above in order to determine the relevance of a document and to build a comprehensive topic – specific library for the benefit of specific users.

As to dependent claim 9, Chakrabarti teaches the experience with the document during the plurality of previous crawls is derived from historical information associated with the document address specification *(e.g., the preferred web page table 32 includes various administrative fields ...indicating the date and time the web page was last modified; col.5, lines 47-60).*

Response to Argument

5.   Applicants' arguments with respect to claims 1-20 have been considered but they are not

persuasive.

Applicant argues that *Chakrabarti et al. clearly does not disclose determining whether to*

access a document during a current crawl with the aid of a probabilistic model that is based on

the probability that the document has changed since a previous crawl. (Remarks, page 8, lines

11-13)

In response, "determining whether to access a document during a current crawl with the

aid of a probabilistic model that is based on the probability that the document has changed since

a previous crawl" was not previously claimed. The Examiner believes that the added features are

met by Chakrabarti. Note the rejection above.

Conclusion

6.   The prior art made of record and not relied upon is considered pertinent to applicant's

disclosure.

Houser et al.   U.S Patent No. 5,606,609   issued: Feb. 25,1997

Najork et al.   U.S Patent No. 6,263,364   issued: Jul. 17, 2001

Lam et al., "Automatic Document Classification Based on Probabilistic Reasoning:

Model and Performance Analysis", IEEE, 01/1997, pages 2719-2723.

Huang el al., " Design and implementation of a Chinese full-text retrieval system based

on a probabilistic model", IEEE, 10/1993, pages 1090-1093.

7.   Applicant's amendment necessitated the new ground(s) of rejection presented in this

Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a).

Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

        A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within TWO

MONTHS of the mailing date of this final action and the advisory action is not mailed until after

the end of the THREE-MONTH shortened statutory period, then the shortened statutory period

will expire on the date the advisory action is mailed, and any extension fee pursuant to 37

CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event,

however, will the statutory period for reply expire later than SIX MONTHS from the mailing

date of this final action.

8.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Maikhanh Nguyen whose telephone number is (703) 306-0092.

The examiner can normally be reached on Monday - Friday from 9:00am – 5:30 pm.

        If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Joseph H Feild can be reached on (703) 305-9792.

        The fax phone number for the organization where this application or proceeding is

assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent

Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished

applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Maikhanh Nguyen
July 20, 2004

JOSEPH FEILD
SUPERVISORY PATENT EXAMINER